



Игорь Гайдышев

Решение научных и инженерных задач средствами Excel, VBA и C/C++



+ CD-ROM

- Обработка и анализ данных
- Статистический контроль качества
- Математическое моделирование
- Распознавание образов
- Создание макросов XLM и VBA
- Разработка DLL и XLL для MS Excel



МАСТЕР РЕШЕНИЙ

Игорь Гайдышев

**Решение научных
и инженерных задач
средствами Excel,
VBA и C/C++**

Санкт-Петербург
«БХВ-Петербург»
2004

УДК 681.3.068+800VBA,C++
ББК 32.973.26-018.1
Г14

Гайдышев И. П.

Г14 Решение научных и инженерных задач средствами Excel, VBA и C/C++. — СПб.: БХВ-Петербург, 2004. — 512 с.: ил.

ISBN 5-94157-477-0

Книга посвящена практическим вопросам программирования. Подробно рассматривается технология создания недорогого и максимально адаптированного для пользователей всех квалификаций программного обеспечения для анализа данных и математического моделирования. Изложение ведется как в приложении к официальным, так и — впервые в литературе — альтернативным средствам разработки, причем последним уделяется большое внимание. На прилагаемом компакт-диске даются полные исходные тексты программ, проекты для различных средств разработки и дистрибутивы программ и дополнительных модулей для электронных таблиц. Все программы тщательно протестированы. Книга будет полезна в качестве практического пособия преподавателям и студентам, программистам, научным работникам и инженерам-исследователям.

Для программистов

УДК 681.3.068+800VBA,C++
ББК 32.973.26-018.1

Группа подготовки издания:

| | |
|-------------------------|----------------------------|
| Главный редактор | <i>Екатерина Кондукова</i> |
| Зам. главного редактора | <i>Игорь Шишигин</i> |
| Зав. редакцией | <i>Григорий Добин</i> |
| Редактор | <i>Елена Кашлакова</i> |
| Компьютерная верстка | <i>Татьяны Валерьяновы</i> |
| Корректор | <i>Евгений Камский</i> |
| Оформление серии | <i>Via Design</i> |
| Дизайн обложки | <i>Игоря Цырульниковы</i> |
| Зав. производством | <i>Николай Тверских</i> |

Лицензия ИД № 02429 от 24.07.00. Подписано в печать 19.03.04.
Формат 70×100^{1/16}. Печать офсетная. Усл. печ. л. 41,28.
Тираж 3000 экз. Заказ №
"БХВ-Петербург", 190005, Санкт-Петербург, Измайловский пр., 29.

Гигиеническое заключение на продукцию, товар № 77.99.02.953.Д.001537.03.02 от 13.03.2002 г. выдано Департаментом ГСЭН Минздрава России.

Отпечатано с готовых диапозитивов
в Академической типографии "Наука" РАН
199034, Санкт-Петербург, 9 линия, 12.

Содержание

| | |
|---|------------|
| От автора | 7 |
| Введение | 8 |
| Глава 1. Разработка элементов пользовательского интерфейса на языке VBA..... | 18 |
| Решаемая проблема | 18 |
| Редактор Visual Basic | 20 |
| Создание меню средствами VBA | 22 |
| Создание диалоговых панелей | 29 |
| Ввод, верификация данных средствами VBA и обработка аварийных ситуаций .. | 38 |
| Рекомендуемая литература | 48 |
| Глава 2. Разработка дополнительных модулей для Microsoft Excel..... | 49 |
| Решаемая проблема | 50 |
| Разработка надстроек для Microsoft Excel..... | 50 |
| Дополнительные модули..... | 58 |
| Стандартные DLL..... | 65 |
| Организация работы Microsoft Excel с программами и пользовательскими функциями | 68 |
| Особенности отладки и тестирования расчетных программ..... | 72 |
| Ресурсы Интернет..... | 75 |
| Рекомендуемая литература | 76 |
| Глава 3. Предварительная обработка данных на VBA и C/C++ | 78 |
| Решаемая проблема..... | 78 |
| Проверка типа распределения эмпирических данных | 79 |
| Обработка пропущенных данных | 111 |
| Обработка выбросов..... | 118 |
| Руководство пользователя..... | 127 |
| Тестирование и пример применения | 135 |
| Ресурсы Интернет..... | 136 |
| Рекомендуемая литература | 136 |
| Глава 4. Информационный анализ на «чистом» VBA | 138 |
| Решаемая проблема..... | 138 |
| Особенности решения..... | 138 |

| | |
|---|------------|
| Научные алгоритмы | 139 |
| Руководство пользователя и сообщения об ошибках | 144 |
| Рекомендуемая литература | 147 |
| Глава 5. Язык макросов Microsoft Excel и VBA: | |
| статистическая проверка гипотез | 148 |
| Решаемая проблема | 148 |
| Особенности решения | 149 |
| Научные алгоритмы | 150 |
| Руководство пользователя | 173 |
| Пример применения | 177 |
| Рекомендуемая литература | 178 |
| Глава 6. Многомерный анализ: факторный анализ в VBA и С | 180 |
| Решаемая проблема | 180 |
| Особенности решения | 182 |
| Научные алгоритмы | 183 |
| Руководство пользователя | 208 |
| Дальнейшие исследования | 211 |
| Рекомендуемая литература | 211 |
| Глава 7. Распознавание образов на VBA и С | 213 |
| Решаемая проблема | 213 |
| Распознавание образов без обучения | 214 |
| Распознавание образов с обучением | 245 |
| Пример применения | 296 |
| Рекомендуемая литература | 296 |
| Глава 8. Корреляционный анализ в кодах VBA и С | 298 |
| Решаемая проблема | 298 |
| Особенности решения | 299 |
| Научные алгоритмы | 299 |
| Руководство пользователя | 309 |
| Дальнейшие исследования | 312 |
| Пример практического применения | 312 |
| Рекомендуемая литература | 313 |
| Глава 9. Методы математического моделирования в программах VBA и С/С++ ... | 315 |
| Решаемая проблема | 315 |
| Особенности решения | 317 |
| Научные алгоритмы | 318 |
| Численное решение дифференциальных уравнений | 331 |

| | |
|---|------------|
| Идентификация параметров дифференциальных уравнений..... | 367 |
| О приложении дифференциальных уравнений к вычислению статистических распределений..... | 370 |
| Ошибки численных схем..... | 372 |
| Руководство пользователя и пример применения..... | 373 |
| Дальнейшие исследования..... | 377 |
| Рекомендуемая литература..... | 377 |
| Глава 10. Рекурсии на примере многочленов в виде XLL на XLM и C..... | 380 |
| Решаемая проблема..... | 380 |
| Особенности решения..... | 381 |
| Научные алгоритмы..... | 381 |
| Руководство пользователя..... | 389 |
| Рекомендуемая литература..... | 390 |
| Глава 11. Статистический контроль качества и графические средства VBA..... | 392 |
| Решаемая проблема..... | 392 |
| Научные алгоритмы..... | 393 |
| Руководство пользователя..... | 403 |
| Ресурсы Интернет..... | 405 |
| Рекомендуемая литература..... | 405 |
| Глава 12. Разработка справочной системы и генерация дистрибутива..... | 406 |
| Решаемая проблема..... | 406 |
| Разработка справочной системы..... | 407 |
| Генерация дистрибутива..... | 426 |
| Рекомендуемая литература..... | 434 |
| Приложение 1. Разработка дополнительных модулей для StarOffice и OpenOffice.org..... | 435 |
| Решаемая проблема..... | 435 |
| История StarOffice..... | 436 |
| Средства разработки для StarCalc/Calc..... | 437 |
| Особенности разработки дополнительных модулей для StarCalc/Calc..... | 438 |
| Использование add-in для StarCalc/Calc..... | 444 |
| Рекомендуемая литература..... | 445 |
| Приложение 2. Статистические распределения..... | 446 |
| Вычисление распределений..... | 446 |
| Нормальное распределение..... | 456 |
| Хи-квадрат - распределение..... | 458 |

| | |
|--|------------|
| Распределение Стьюдента | 459 |
| Рекомендуемая литература | 461 |
| Приложение 3. Разработка консольных приложений на C/C++ | 463 |
| Решаемая проблема | 463 |
| Рекомендуемая литература | 465 |
| Приложение 4. Средства программирования (обзор)..... | 466 |
| Операционные системы..... | 467 |
| Языки научного программирования..... | 481 |
| Ресурсы Интернет | 489 |
| Рекомендуемая литература | 489 |
| Приложение 5. Работа с прилагаемым компакт-дискom | 491 |
| Технические требования | 491 |
| Установка программного обеспечения..... | 491 |
| Запуск программ и работа с ними..... | 496 |
| Техническая поддержка | 496 |
| Правовые вопросы использования программного обеспечения | 496 |
| Содержимое прилагаемого компакт-диска..... | 501 |
| Заключение | 502 |

ОТ АВТОРА

Первая наша книга восполнила только половину широкого зазора между теорией и практикой научного анализа данных и математического моделирования — она представила современные, надежные и быстрые реализации проверенных алгоритмов. На ее создание, вместе с программной отработкой алгоритмов и практическими исследованиями, ушло полтора десятка лет. Благодаря длительному тестированию реализация программных продуктов занимает, как это и планировалось, минимальное время на базе любой программной платформы, поддерживающей применяемые нами средства разработки на языках C и C++.

Нынешняя книга, благодаря качественной предварительной работе, создавалась гораздо быстрее. Она завершает проект превращения формул в работающие программы. Мы постарались сделать это прямым и понятным путем. Все рассмотренные в книге программы работают и уже приносят практическую выгоду тысячам пользователей. После завершения работы мы, с некоторым удивлением для себя, поняли, что, по всей видимости, нам повезло найти оптимальное решение по доставке научных алгоритмов анализа массовому, в первую очередь, отечественному пользователю.

Наш молодой коллега однажды спросил, ознакомившись с версткой книги: «Неужели это все из головы?» Это из головы, хотя и базируется на исследованиях, сделанных и другими авторами, и нами. Так было принято в науке: по крупицам собирать факты, делать выводы и, в конце концов, становиться базой для дальнейших исследований других авторов. В нашей книге читатель найдет только результаты оригинальных исследований и полные тексты оригинальных работающих программ. Здесь нет компиляции найденных в сети материалов, которые так любят некоторые студенты - представители поколения Интернета — предъявлять в качестве рефератов и курсовых работ. Здесь нет заимствований размером в полглавы и, надеемся, нет слишком уж грубых ошибок.

Надеемся, читатель благожелательно оценит наш труд.

Игорь Гайдышев



Введение

«Компьютерный мир захвачен потребительским образом мышления, для которого важна не культурная ценность той или иной вещи, а лишь то, сколько эта вещь стоит».

Т. Полгар

«Когда машина станет универсальной, что будут делать люди?»

К. Ривз

Проблема

В нашей книге изложение практики программирования мы ведем в применении к научно-техническим программам, что совсем не исключает применение рассмотренных алгоритмов для любых других прикладных задач. Методики с успехом применялись, кроме решения научных и инженерных задач, при разработке различных программ экономического и хозяйственного назначения. Это было доказано нашими коллегами по работе в Российском научном центре «Восстановительная травматология и ортопедия» имени академика Г.А. Илизарова (Генеральный директор профессор В.И. Шевцов).

Итак, к научно-техническим программам можно отнести:

- программы математического моделирования;
- программы статистической обработки экспериментальной информации, анализа данных различной природы; программы статистического контроля качества;
- системы распознавания образов;
- программы аналитических вычислений;
- экспертные системы;
- программы управления и автоматизации эксперимента и производственных процессов; программы управления различными устройствами как гражданского, так и военного назначения¹.

Эти программы в реальной жизни могут исполняться на различных типах компьютеров, существовать в виде бинарных модулей (software) или быть

¹ По словам Уинстона Черчилля, «что бы ученые ни делали, у них всегда получается оружие...».

встроенными в различные устройства, а также пересекаться. Программы последнего типа часто являются заказными и имеют узкое специальное применение, не позволяющее не только описать их в доступной литературе, но даже предположить об их существовании. В нашей книге мы остановимся на создании программ, в основном относящихся к первым трем разделам научно-технического программирования.

Несмотря на обширную номенклатуру программ математического моделирования, анализа данных и распознавания образов, слишком оптимистично утверждать, что это программное обеспечение распространено в России настолько широко, насколько это необходимо для обеспечения статистической обработки данных, анализа, прогнозирования и математического моделирования. Можно предположить, что основные причины сложившейся ситуации:

- ❑ Высокая стоимость программ. Научная сложность, продолжительность процесса разработки и отладки прикладных программных продуктов рассматриваемого класса и малый тираж никогда не смогут снизить их цену до приемлемых для массового потребителя величин.
- ❑ Необходимость изучения новых программных продуктов. Некоторые программы имеют нестандартные или слишком запутанные интерфейсы, и выполнить нужную функцию, не уделив длительного времени обучению, нелегко даже специалисту, профессионально разбирающемуся в методах анализа.

Положение несколько сглаживается распространением так называемых «пиратских» копий программ. Отсутствие гарантий правильной работы программы², документации и технической поддержки, юридическое преследование – только некоторые из бед, подстерегающих пользователя на этом опасном пути. Противоправные действия по использованию «пиратского» программного обеспечения пользователи пытаются оправдать благородством целей. Но согласно рекомендациям по представлению результатов научных изысканий в печати, в любой научной работе следует указать, какими программными средствами научного анализа пользовался автор. Ссылка на «пиратскую» копию программы будет нонсенсом настолько же, насколько сообщение о том, что научное исследование выполнено с помощью украденного прибора.

Назрела необходимость создания максимально доступного пользователям разной квалификации программного анализа данных. Доступные для изучения, тщательно прокомментированные, простые и понятные тексты программ дают возможность проверить правильность реализации алгоритмов и информационную безопасность программного обеспечения. Публикация алгоритмов и кодов

² О соотношении принципа «As is» и прав покупателя получить за свои деньги качественный программный продукт см. статью Г. Григоренко в «КомпьютерПресс», № 1, 2003 г., с. 179.

в книге позволяет сделать это, не подвергая опасности свою интеллектуальную собственность. Не основной, но как надеется автор, важной целью книги, является демонстрация сложности решения вычислительных задач, не пугающей, а вполне преодолимой.

Предмет книги

Программный анализ данных может быть выполнен несколькими способами:

- ❑ Stand-alone программа (package) или приложение (application). Для анализа данных с помощью рассматриваемой категории программ, в принципе, не требуется другое программное обеспечение, кроме операционной системы.
- ❑ Web-based решения: апплет (applet) или сценарий (script). Выполнение производится под управлением браузера (browser), независимо от аппаратно-программной платформы.
- ❑ Надстройка (дополнительный модуль, add-in) к другой программе. Нужно иметь базовую программу, позволяющую работать с надстройками (это может быть текстовый процессор, электронные таблицы, программа обработки векторной или растровой графики, программа анализа данных и другие программы), а надстройка призвана увеличить ее функциональность.

Книга посвящена надстройкам. Рассмотрены разнообразные возможности для разработки модулей анализа и математического моделирования, значительно усиливающих аналитические способности электронных таблиц. Не говорится об инструментариим программиста и библиотеках разработчика, главным вопросом при использовании которых является решение дилеммы: применить их на свой страх и риск (в том числе финансовый) или создать свои собственные библиотеки, что не так уж и сложно. Не рассматриваются также проблемы анализа данных на примерах известных программных продуктов.

В соответствии с нашими представлениями, в книге рассматривается процесс разработки программы как наукоемкого изделия, от строгого теоретического обоснования до практически готового программного продукта. При желании, читатель может пройти весь этот путь либо подключиться к решению на любом этапе.

Популярные электронные таблицы Microsoft Excel обладают набором тщательно протестированных статистических функций (параметрические тесты, стандартные распределения, регрессионный анализ). Пакет анализа Excel содержит такие методы, как:

- ❑ описательная статистика,
- ❑ параметрические методы проверки гипотез,

- дисперсионный анализ,
- вычисление коэффициента корреляции.

Не хватает, как минимум, следующих методов:

- проверка типа эмпирического распределения, в частности, проверка нормальности распределения,
- обработка пропущенных значений,
- обработка выбросов,
- информационный анализ,
- непараметрические методы проверки гипотез,
- факторный анализ,
- распознавание образов без обучения и с обучением,
- корреляционный анализ,
- статистический контроль качества.

В нашей книге в качестве примеров рассмотрены проекты, реализующие все перечисленные методы анализа Microsoft Excel, а в *приложении 1* приводится пример альтернативного решения для электронных таблиц офисных пакетов OpenOffice.org и Sun StarOffice. Большинство программных решений представляют собой модули разработанных нами программных продуктов AtteStat и ME.com. Часть модулей распространяется на коммерческой основе, часть свободно распространяется. Исходные тексты всех модулей доступны.

В показанных листингах применяется ряд функций, обычно без каких-либо модернизаций, улучшений и исправлений, на языках C и C++ (операции с массивами, линейная алгебра, статистические распределения) из нашей первой книги, но исходные тексты даны в приложении на компакт-диске. Если было удобно для реализации конкретных программ, когда применение языка C казалось излишним, составлялись аналоги на языке VBA.

Структура книги

Книга состоит из двенадцати глав и пяти приложений.

Глава 1 посвящена избранным вопросам создания элементов пользовательского интерфейса на языке VBA. Глава не может служить справочником и руководством по VBA. В главе описаны только минимальные сведения, которых достаточно, чтобы создать работающий интерфейс к модулям программы анализа данных, представленным в главах 3–8. Автор работает с VBA не очень продолжительное время, однако мощностю и простотой данного языка программирования позволила добиться устойчивых успехов.

Глава 2 посвящена вопросам программирования дополнительных модулей для Microsoft Excel. В ней нет теоретических сведений. Однако в ней нет и лишних сведений, втягивающих в обсуждение сложных вопросов системного программирования и, как огромный рекламный щит, отвлекающих от прямой дороги быстрого практического создания нормально функционирующей программы. Приведенная информация тщательно проверена и автор полагает, что все предлагаемые решения действительно работают. Техническая информация, без сомнения, обратит внимание программистов на удобство и простоту создания дополнительных модулей для Microsoft Excel (да и других программ) и повернет многих программистов, которых отпугнули неполные материалы источников и электронной документации, к данной технологии. На разработку материалов главы было потрачено определенное время, т. к. ряд ключевых сведений в литературе и в сопроводительной программной документации совершенно не представлен. Автор надеется, что его труд по исследованию не пропал даром, и многим программистам сэкономит немало драгоценного времени.

Главы 3–8 посвящены различным особенностям создания научных программ для Microsoft Excel. В качестве примеров в главах рассмотрены реально работающие проекты конкретных разработок по статистическому анализу данных. Рассмотренные подходы могут быть использованы не только для научного программирования, но и для решения любых других прикладных задач. При написании глав использован многолетний опыт работы автора по составлению и применению в практической деятельности различных статистических программ.

Глава 9 посвящена одному из важнейших разделов математического моделирования и научного программирования — решению обыкновенных дифференциальных уравнений. Просто и естественно, но в то же время научно строго, обращается внимание на важнейшие проблемы, часто оставляемые без внимания авторами популярных практических руководств. При написании главы использован опыт работы автора с математическими моделями сложных инженерных механических систем.

Глава 10 посвящена разработке рекуррентных алгоритмов на примере программирования алгоритмов вычисления многочленов.

Глава 11 посвящена наиболее яркому практическому применению элементарных методов анализа данных — некоторым методам статистического контроля качества.

Глава 12 посвящена разработке справочной системы программы и генерации дистрибутива (установочного пакета). Рассмотрены наиболее современные, на момент написания книги, и удобные, с точки зрения автора, средства для решения данных задач.

Приложение 1 посвящено решениям по анализу данных для альтернативных электронных таблиц. Дан функциональный аналог решения из главы 5, но

для электронных таблиц других типов. Приводятся уникальные сведения о разработке дополнительных модулей, не затронутые в литературе и отсутствующие в документации.

В приложении 2 рассмотрены стандартные статистические распределения, применяемые в книге.

В приложении 3 рассмотрены вопросы создания консольных приложений, необходимых для отладки алгоритмов.

В приложении 4 дан краткий обзор средств программирования с интересным фактическим материалом.

В приложении 5 подробно описана работа с прилагаемым программным обеспечением. Несмотря на кажущуюся тривиальность, представленные сведения могут оказаться полезными и начинающему пользователю компьютера, и опытному программисту.

Каждая глава заканчивается списком литературы, причем ссылки на источники в тексте глав не делаются, как не делаются и какие-либо объемные заимствования из источников. Это сделано намеренно. Приводимой в книге информации достаточно для самостоятельной работы с прилагаемыми методами расчета и программным обеспечением. Если же приведенной в книге информации все же покажется мало, читатель всегда может собрать набор литературы, рекомендуемый нами, и изучить проблему более подробно. Часть источников давно стала библиографической редкостью, поэтому автор особенно старался сделать ссылки на книги, изданные в последние годы.

Содержание прилагаемого компакт-диска

Неотъемлемой частью книги является компакт-диск, который содержит полные тексты и проекты всех модулей программного обеспечения, описанного в книге. Благодаря наличию законченных проектов книгу в комплекте с компакт-диск можно логическим продолжением первой книги автора.

На компакт-диске находятся также дистрибутивы всех описанных в книге модулей учебных версий программного обеспечения анализа данных и математического моделирования. Отличия учебной версии от полной коммерческой версии заключаются только в отсутствии справочной системы и печатного руководства пользователя, поставляемого по специальному заказу. Учебные версии поставляются только вместе с книгой и сопровождаются исходными кодами (для коммерческих версий они доступны по особому запросу), поэтому книга в комплекте с компакт-диск заменяет собой справочную систему. Работа с прилагаемым компакт-диск описана в приложении 4.

Лицензионное программное обеспечение

Для установки и использования программного обеспечения, дистрибутивы которого содержатся на прилагаемом компакт-диске, необходимы лицензионные программы:

- ❑ Операционная система из семейства Microsoft Windows 95/98/98SE/NT4/2000/XP.
- ❑ Электронные таблицы Microsoft Excel, входящие в пакет офисных приложений Microsoft Office 97, 2000, XP или 2003 любой редакции, которые мы используем в качестве платформы для представленных в книге разработок.
- ❑ Пакет офисных приложений Sun StarOffice 5.2 и выше или пакет офисных приложений с открытыми исходными кодами OpenOffice.org версии 1.0 и выше.

Для работы с книгой в плане изучения, адаптации и совершенствования модулей представленного программного обеспечения дополнительно, в зависимости от принятого решения, может потребоваться следующее лицензионное программное обеспечение:

- ❑ Microsoft Excel 97 Developer's Kit — это компонент официального набора электронной документации MSDN Library от Microsoft Corporation, содержащий почти полную информацию по всем аспектам программирования для Microsoft Excel. Для получения некоторых материалов, которые должны присутствовать в данном пакете, читателю придется посетить сайт MSDN.
- ❑ Система разработки приложений Microsoft Visual C++ 5.0 и выше любой редакции предназначена для разработки проектов из глав 3–10, кроме проекта из приложения 1, для разработки которого пришлось применить Bloodshed Dev-C++ 4.01. Рекомендации по установке и использованию Microsoft Visual C++ см. в главе 2.
- ❑ Система разработки приложений Borland C++ версии 5.5 предназначена для разработки проектов из глав 3–10, как альтернатива дорогостоящему решению Microsoft Visual C++.
- ❑ Система разработки приложений Bloodshed Dev-C++ версии 4.01 и выше³ предназначена для разработки проектов из глав 3–5, 9 и 10, конкурируя с системами разработки Microsoft Visual C++ и Borland C++, и разработки проекта из приложения 1, для которой Microsoft Visual C++ не оказался полезным. Для работы с проектами из глав 6–8 Bloodshed Dev-C++ применить не удалось.

Дополнительные инструментальные средства, необходимые для создания программного продукта, такие, как компиляторы справочной системы или генера-

³ На момент данной публикации все проблемы с версией 5 решены не были.

торы установочных пакетов (дистрибутивов), работа с которыми рассмотрена в главе 12, можно получить в комплекте со средствами разработки, приобрести отдельно или из сети Интернет, обратившись по указанным ссылкам.

Ресурсы Интернет

С большинством программных продуктов по анализу данных всех типов можно ознакомиться на сайте **www.statistics.com**⁴, где даны подробные ссылки на web-based ресурсы, но не делается различий между stand-alone программами и надстройками. Поэтому перечислим наиболее известные научные надстройки для популярных электронных таблиц.

Надстройки, выполняющие научные расчеты в среде электронных таблиц Excel, поставляют: Addinsoft SARL (**www.xlstat.com**), Alpha System Ltd (**www.user.cityline.ru/~alphasys**), AnalyCorp (**www.analycorp.com**), Analyse-it Software Ltd (**www.analyse-it.com**), BaRaN Systems LLC (**www.baran-systems.com**), Digital Computations Inc. (**www.sigmazone.com**), JWalk & Associates Inc. (**j-walk.com/ss**), Numerical Algorithms Group Limited (**www.nag.co.uk**), NeuroDimension Inc. (**www.nd.com**), NeurOK LLC (**www.neurok.ru**), statistiXL (StatistiXL, **www.statistixl.com**), UNISTAT Ltd (**www.unistat.com**), Vector One Corporation Ltd (**www.vstat.km.ru**) и другие.

Вы можете ознакомиться с предлагаемыми решениями и почерпнуть из них некоторые идеи по усовершенствованию номенклатуры методов и интерфейса своих собственных программ, а иногда и посмотреть, как не нужно делать, если целью разработки поставлено создание полезной программы.

Благодарности

Автор искренне благодарит:

- руководителя АСУ областной больницы д.м.н. В.А. Уткина (г. Курган), за полезное обсуждение проблем программной реализации научных алгоритмов статистического анализа данных,
- кандидата наук В.А. Биллига (Тверской госуниверситет), за проявленное внимание и подробное пояснение механизма разработки надстроек к программному обеспечению Microsoft Office (в основном касательно Microsoft Word),
- Али Абубакирова, генерального директора ООО «Физтех-софт», за консультации по истории операционных систем отечественной разработки,

⁴ Здесь и далее все ссылки актуальны на момент составления информации и могут быть без предупреждения изменены.

- студента-химика Курганского государственного университета и системного администратора РНЦ «ВТО» им. академика Г.А. Илизарова, своего сына А.И. Гайдышева, за тестирование разработанного нами программного обеспечения, его апробацию в своей учебной и научной работе и поддержку корпоративной и домашней компьютерных сетей на базе Microsoft Windows 2000 Server и ALT Linux,
- самого увлеченного компьютерами человека, свою жену С.А. Гайдышеву, за терпение, поддержку, тестирование программ и консультации по использованию электронных коммуникаций,
- руководителя рекламного подразделения М.А. Уланову (РНЦ «ВТО», г. Курган), за название прилагаемого к книге программного обеспечения анализа данных AtteStat, как за определенный лингвистический смысл в немецком и русском языках, так и за классическую для статистических программ аббревиатуру,
- старшего эксперта по выставкам Н.А. Березину (РНЦ «ВТО»), за помощь в работе с англо- и франкоязычными материалами, полезное обсуждение применения математических алгоритмов в гуманитарной сфере, предложенный оригинальный девиз к книге.

Кроме того, автор благодарит всех коллег по прошлой и настоящей работе в стенах РНЦ «ВТО», которые явились инициаторами написания того или иного модуля представленного в книге программного обеспечения:

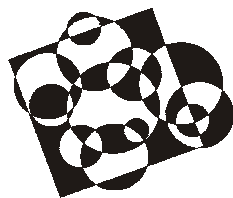
- руководителя отдела телемедицины д.м.н. М.М. Шудло, за предложение создания модуля информационного анализа и тестирование его на своих научно-медицинских экспериментальных данных,
- научного сотрудника к.м.н. А.Н. Ерохина, за идею точного расчета непараметрического критерия Колмогорова–Смирнова (и сделавшего это «вручную») в электронных таблицах Microsoft Excel, с которого началась разработка модуля по непараметрической статистике,
- руководителя подразделения к.м.н. А.Д. Наумова и научного сотрудника к.м.н. Н.А. Гордиевских, за предложение выполнить факторный анализ своих научно-медицинских экспериментальных данных,
- заведующего отделением к.м.н. Д.А. Попкова, за интересную идею сделать кластерный анализ средствами Microsoft Excel.

Автор особо благодарит Российское представительство корпорации Microsoft за предоставленное в комплекте с лучшим на тот момент компьютером R&K Wiener 4 лицензионное программное обеспечение — операционную систему Microsoft Windows XP Home Edition и стандартный выпуск офисного пакета программ Microsoft Office XP. Это позволило протестировать все разработанные программы на совместимость с самым современным системным и прикладным программным обеспечением.

Рекомендуемая литература

1. Bailar J.C., Mosteller F. Guidelines for Statistical Reporting in Articles for Medical Journals. Amplifications and Explanations // *Annals of Internal Medicine*, 15 February 1988, 108:266–273. Available from World Wide Web: < <http://www.acponline.org/journals/resource/guidelines.htm> >.
1. Боровиков В.П. STATISTICA: искусство анализа данных на компьютере. Для профессионалов. — СПб.: Питер, 2001.
2. Бююль А., Цефель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей. — СПб.: ООО «ДиаСофтЮП», 2001.
3. Васильков Ю.В., Василькова Н.Н. Компьютерные технологии вычислений в математическом моделировании. — М.: Финансы и статистика, 2002.
4. Дюк В. Обработка данных на ПК в примерах. — СПб.: Питер, 1997.
5. Кулаичев А.П. Методы и средства анализа данных в среде Windows. STADIA. — М.: Информатика и компьютеры, 1999.
6. Макарова Н.В., Трофимец В.Я. Статистика в Excel: Учебное пособие. — М.: Финансы и статистика, 2002.
7. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере. — М.: ИНФРА-М, 1999.⁵
8. Орвис В. Дж. Excel для ученых, инженеров и студентов. — Киев: Юниор, 1999.

⁵ В 2003 г. выпущено дополненное переиздание.



ГЛАВА 1

Разработка элементов пользовательского интерфейса на языке VBA

«Поскольку сегодняшний пользователь ЭВМ обычно имеет профессиональные интересы, выходящие за рамки программирования, то пользовательский интерфейс должен стимулировать эти интересы, а не сдерживать их».

А. Стивенс

«У Вас есть выбор, на каком языке писать ту или иную часть приложения. Так, пользовательский интерфейс приложения Вы, скорее всего, будете создавать на Microsoft Visual Basic, но прикладную логику лучше всего реализовать на C++.»

Дж. Рихтер

Для работы с материалами главы понадобятся лицензионные средства разработки и тестирования программ:

- Microsoft Excel 97 или выше;
- Microsoft Excel Developer's Kit версии 8 или выше.

Потребуется знание VBA.

Решаемая проблема

Современная программа, предназначенная для обработки научных, финансовых, коммерческих или иных данных, предоставляет возможности для

хранения массивов числовой информации (учитывая возможность преобразования в числовой других видов информации) и программные средства для их обработки. Такой концепции наиболее полно отвечают электронные таблицы, оснащенные дополнительными аналитическими инструментами. Без потери общности, мы можем говорить о научных программах, потому что результаты наших исследований применимы для решения широкого круга задач.

Лучшие специализированные программы статистического анализа данных:

- ❑ NCSS — Number Cruncher Statistical Systems (NCSS, www.ncss.com),
- ❑ SPSS (SPSS Inc., www.spss.com),
- ❑ STADIA (НПО «Информатика и компьютеры», statsoft.msu.ru)
- ❑ STATISTICA (StatSoft Inc., www.statsoft.com),
- ❑ SYSTAT (SYSTAT Software Inc., www.systat.com)

и другие — построены также на концепции электронных таблиц. Однако средства манипуляции данными в перечисленных программах несравнимо более ограничены, чем в любых электронных таблицах. Это и понятно — производителям лучших программ анализа данных вряд ли по силам конкурировать в части интерфейса с производителями любых электронных таблиц. У этих программных продуктов — разные задачи и группы пользователей, пересекающиеся незначительно. Однако мы покажем, что вполне возможно создать программный продукт, сосредоточившись на реализации алгоритмов анализа данных и воспользовавшись всеми доступными средствами, предлагаемыми лучшими в мире электронными таблицами.

Остановимся кратко на истории электронных таблиц производства Microsoft, с ними нам предстоит работать на протяжении всей книги. Под руководством Ч. Саймони (Charles Simonyi), перешедшим в Microsoft из исследовательского центра Xerox PARC, в 1981 г. были разработаны электронные таблицы Multiplan для операционной системы MS-DOS, как, годом позднее, и текстовый процессор Word для MS-DOS. Программный продукт Microsoft Multiplan для Apple Macintosh был анонсирован в 1984 г. Программа Multiplan была написана на языке программирования C, чем была обеспечена ее легкая переносимость на микрокомпьютеры с различной архитектурой. Однако эти электронные таблицы оказались менее совершенными, чем продукты конкурентов.

Электронные таблицы (иначе, табличный процессор) Microsoft Excel были разработаны в 1985 г. для платформы Macintosh, и это было стратегически правильным решением. Графические возможности Apple Macintosh и Microsoft Excel обеспечили успех и тому, и другому продукту. В 1987 г. появилась версия Microsoft Excel 2 для операционной системы Microsoft Windows.

Электронные таблицы Microsoft Excel 4 были впервые оснащены специальной версией языка программирования Visual Basic Applications Edition.

Первой 32-разрядной версией стали электронные таблицы Microsoft Excel 7 (другое название Microsoft Excel 95), выпущенные в 1995 г. Версия 8 (Microsoft Excel 97) была разработана в 1997 г. и получила новый интерфейс Visual Basic for Application (VBA), единый для всех компонентов Microsoft Office 97. Благодаря Microsoft Excel и VBA, практически на каждом современном компьютере в распоряжении пользователя имеется мощная система программирования. Компонентом пакета офисных приложений Microsoft Office 2000 стала версия 9 (Microsoft Excel 2000), Microsoft Office XP — версия 10 (Microsoft Excel 2002), Microsoft Office 2003 — версия 11 (Microsoft Excel 2003). Пакет Microsoft Excel доступен и в качестве отдельного продукта для Microsoft Windows и Apple Macintosh.

Представленные в книге дополнительные модули для электронных таблиц Microsoft Excel корректно работают в локализованных версиях Microsoft Excel 97, 2000, 2002 и 2003.

Мы хотим, чтобы с помощью сведений данной главы читатель любой квалификации мог разобраться (или использовать как готовый рецепт, не вникая в подробности) в VBA и легко, как и предполагает VBA, создавать программные проекты любой сложности. Здесь представлен надежно функционирующий каркас полезного приложения, который можно модифицировать или наращивать, по своему усмотрению.

Вывод графиков средствами VBA в настоящей главе не упоминается. Он отдельно рассмотрен в *главе 11*.

Редактор Visual Basic

Разработка программы на языке VBA начинается одинаково для всех представленных в книге проектов, использующих VBA. Все примеры приводятся для русских версий Excel из комплекта офисных приложений Microsoft Office 2000 Professional, под управлением операционной системы MS Windows 98 Second Edition¹, или из комплекта офисных приложений Microsoft Office XP Standard под управлением операционной системы MS Windows XP Home Edition.

После запуска электронных таблиц Microsoft Excel, нажмите комбинацию клавиш <Alt>+<F11> либо выберите из меню Excel последовательность пунктов **Сервис** ▶ **Макрос** ▶ **Редактор Visual Basic**. Появится окно интегрированной среды разработки приложений Visual Basic for Application, как на рис. 1.1.

¹ О типах операционных систем см. приложение 4.

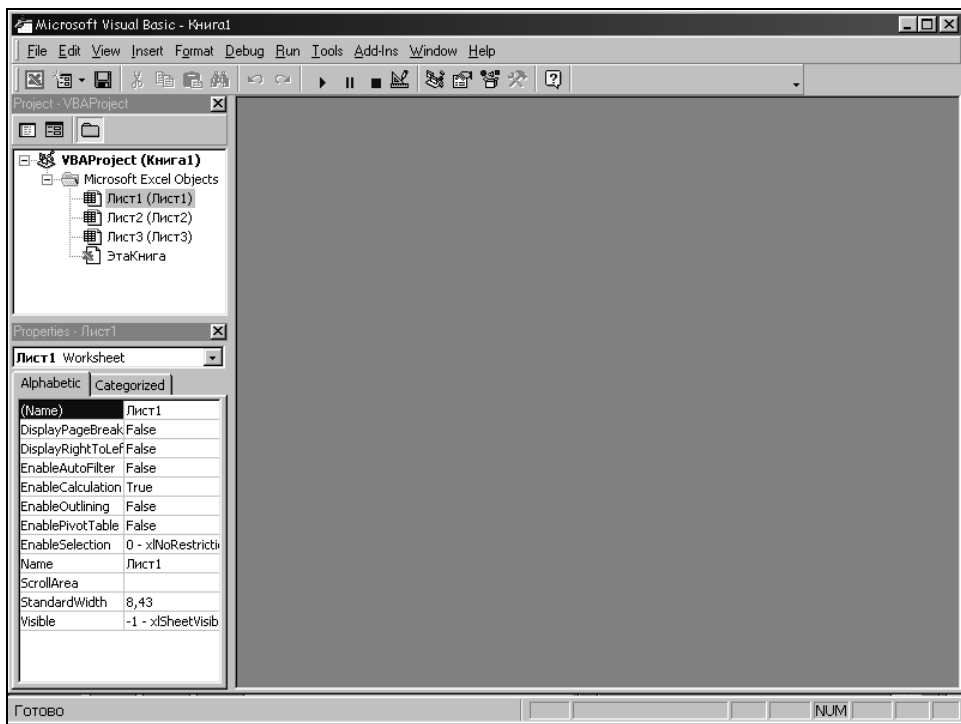


Рис. 1.1. Интегрированная среда разработки приложений Visual Basic for Application

Нажмите правую кнопку мыши в любом месте окна **Project** и выберите из появившегося меню **Insert ► Module**. К проекту будет добавлен новый модуль **Module1** (потом можно переименовать), а в появившееся окно модуля введите исходные тексты программы (макроса).

Следует предусмотреть систему меню для легкого доступа к прикладным алгоритмам и диалоговые панели (формы) для выбора или ввода различных параметров, управляющих работой этих прикладных алгоритмов. В *главах 3* и *5* дополнительно приводятся столь же удобные, хотя и функционально ограниченные, решения, позволяющие обойтись не только без отдельной системы меню, но иногда и вообще без кода VBA, воспользовавшись языком макросов Microsoft Excel.

Чтобы придать нашим разработкам вид серьезных программ, в дальнейшем, мы предполагаем, что разработанные программы будут называться AtteStat и ME.com. Разработка будет модульной, подобно настоящей программе анализа данных. Это, на самом деле, недалеко от истины — после добавления всех модулей Microsoft Excel «превратится» в настоящую программу анализа данных.

Создание меню средствами VBA

Добавление пользовательских меню, обеспечивающих удобный доступ к прикладным алгоритмам модульной программы анализа данных, работающей в среде электронных таблиц Microsoft Excel, средствами VBA не является трудной задачей. Язык визуального программирования VBA как и любой современный развитый язык программирования, предлагает несколько решений. На рис. 1.2 показано, что, примерно, мы хотим получить.

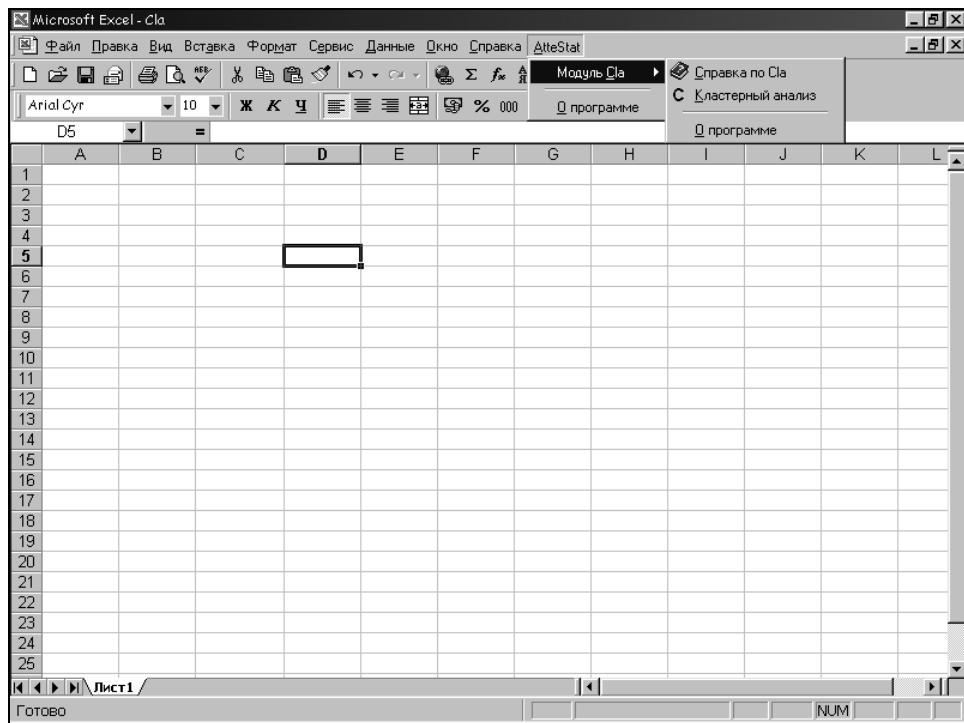


Рис. 1.2. Рабочая книга с пользовательским меню

Наша программа должна добавлять пункт меню к существующему меню Microsoft Excel. Так как программа модульная, будет хорошо, если каждый установленный модуль станет добавлять свое меню к общему меню нашей программы. Это сэкономит место на экране и обеспечит удобный доступ к функциям каждого из установленных модулей.

Программные манипуляции с пользовательскими меню удобно делать в функции `Auto_Open`. Имя функции `Auto_Open` зарезервировано Microsoft Excel в качестве стандартного имени. Это значит, что при загрузке рабочей книги или шаблона, содержащих эту функцию, Microsoft Excel автоматиче-

ски исполняет ее код. В конце главы даются рекомендации по управлению выполнением, в обход любого установленного уровня безопасности Microsoft Excel, что вызвано не пренебрежением к информационной безопасности, а обеспечением требуемой функциональности прикладной программы.

Объекты меню

На листинге показана начальная часть функции `Auto_Open` с описаниями компонентов меню, общих для всех модулей программы `AtteStat`.

Листинг 1.1

```
Sub Auto_Open ()
    '
    ' Загрузка меню
    '
    Dim WorksheetMenuBar As CommandBar    ' Основная панель
    Dim CstmCtrl As CommandBarControl    ' Пункт основного меню
    Dim AtteStatMenu As CommandBarPopup  ' Выпадающее меню AtteStat
    Dim AtteStat As CommandBarControl    ' Пункт меню AtteStat
    Dim MainControl As CommandBarControl ' Пункт меню модуля
    Dim Exist As Boolean                  ' Флаг существования AtteStat
```

Для построения меню используются следующие стандартные средства Microsoft Office из коллекции `CommandBarControls`:

- объект `CommandBar` — панель меню (панель команд);
- объект `CommandBarPopup` — панель выпадающего меню;
- объекты класса `CommandBarControl` — отдельные пункты меню.

Нам необходимо создать общее меню для всех модулей нашей прикладной программы, затем добавить к нему меню конкретного модуля. Идея достаточно проста: каждый модуль, если он загружается первым, создает пункт меню **AtteStat** (или **ME.com**) и добавляет свое собственное меню. Если он загружается не первым, то пункт меню **AtteStat** уже существует, его создавать не нужно, а только добавить меню модуля к существующему меню **AtteStat**.

Поэтому, перед добавлением нового меню модуля, следует решить важную задачу. Необходимо проверить, существует ли в меню Excel пункт меню **AtteStat**, который мы собираемся добавить, так как Microsoft Excel допускает существование в меню двух и более одинаковых пунктов. Для проверки существования

конкретного пункта меню как раз предназначен флаг `Exist` (см. листинг 1.1), а сама проверка выполняется, как показано на листинге 1.2.

Листинг 1.2

```
'
' Проверка наличия меню AtteStat
'
Set WorksheetMenuBar = CommandBars.ActiveMenuBar
Exist = False
For Each CstmCtrl In WorksheetMenuBar.Controls
    If CstmCtrl.Caption = "&AtteStat" Then
        Exist = True
        Set AtteStat = CstmCtrl
    Exit For
End If
Next CstmCtrl
```

Данный фрагмент программы установит (поднимет) флаг `Exist` в положение `True`, если пункт меню `&AtteStat` уже существует (знак `&` обозначает последующий символ «горячей» клавиши, если таковые используются в меню программы; использование знака `&` в поиске пункта меню обязательно). Для нас это будет означать, что создавать новый пункт меню не нужно, а следует добавить пункт загружаемого в данный момент модуля в уже существующее меню. Если флаг `Exist` оставлен в положении по умолчанию `False`, то создадим сначала новое меню **AtteStat**. Свойства меню описаны в следующем разделе.

Как принято в современном программировании, минимальный набор пунктов меню должен содержать, помимо вызова прикладной части программы, также вызов справочной системы и информационного окна (рис. 1.3), содержащего сведения о названии программы, номере версии, обладателе авторских прав и юридических последствиях их коварного нарушения.



Рис. 1.3. Информационное окно интегратора модульной программы ME.com

Свойства меню

На листинге 1.3 показан фрагмент программы, формирующий меню модуля кластерного анализа, логика которого описана в *главе 7*, в зависимости от того, существует ли меню AtteStat на момент исполнения данного кода. Содержательные комментарии объясняют суть выполняемых операторов.

Листинг 1.3

```
'  
' Формирование меню модуля в составе AtteStat  
'  
Dim ClaMenuBar As CommandBar      ' Панель меню модуля  
Dim ClaMenu As CommandBarPopup    ' Выпадающее меню модуля  
Dim ClaControl As CommandBarControl ' Пункт меню модуля  
If Not Exist Then  
    '  
    ' Формирование нового меню AtteStat и добавление к нему меню модуля  
    '  
    Set AtteStatMenu =  
WorksheetMenuBar.Controls.Add (Type:=msoControlPopup, _  
                                Temporary:=True)  
    AtteStatMenu.Caption = "&AtteStat"  
    Set ClaMenu = AtteStatMenu.Controls.Add (Type:=msoControlPopup, _  
                                              Temporary:=True)  
  
    Set MainControl = AtteStatMenu.Controls.Add (Type:=msoControlButton)  
    MainControl.BeginGroup = True  
    MainControl.Caption = "&O программе"  
    MainControl.OnAction = "Cla.MakeAboutMain"  
Else  
    '  
    ' Добавление меню модуля к существующему меню AtteStat  
    '  
    Set ClaMenu = AtteStat.Controls.Add (Type:=msoControlPopup, _  
                                         Temporary:=True, Before:=1)  
  
End If  
ClaMenu.Caption = "Модуль &Cla"
```

```

'
' Формирование меню модуля
'
Set ClaControl = ClaMenu.Controls.Add(Type:=msoControlButton)
ClaControl.FaceId = 983
ClaControl.Caption = "&Справка по Cla"
ClaControl.OnAction = "Cla.MakeHelp"

Set ClaControl = ClaMenu.Controls.Add(Type:=msoControlButton)
ClaControl.FaceId = 82
ClaControl.Caption = "&Кластерный анализ"
ClaControl.OnAction = "Cla.RunSolution"

Set ClaControl = ClaMenu.Controls.Add(Type:=msoControlButton)
ClaControl.BeginGroup = True
ClaControl.Caption = "&О программе"
ClaControl.OnAction = "Cla.MakeAbout"
End Sub

```

Подробные описания методов и свойств даны во встроенной документации VBA и в указанных в конце главы источниках. В данном фрагменте используются свойства:

- ❑ `FaceId` — номер изображения, появляющегося перед пунктом меню;
- ❑ `Caption` — заголовок пункта меню;
- ❑ `OnAction` — метод, исполняемый при выборе пункта меню;
- ❑ `BeginGroup` — разделитель.

Сделаем подробные пояснения относительно свойства `FaceId`. Полный список изображений и соответствующих им номеров не является одинаковым для каждого компьютера, а зависит от установленных на нем программ. Соответствие номера конкретному изображению поможет получить показанная на листинге 1.4 компактная утилита (вспомогательная программа) в коде VBA. Макрос создает два пункта меню (можно изменить их количество, по своему усмотрению), которые в листинге 1.4. мы назвали **1–300** и **301–600** (в примере на компакт-диске пунктов меню немного больше). В каждом пункте меню выводится по 300 изображений с номерами, которые им соответствуют. Программисту остается выбрать понравившееся изображение и указать его номер в свойстве `FaceId`.

Листинг 1.4

```
Sub Auto_Open()  
    Dim TestIDMenuBar As Object  
    Dim TestIDMenu As Object  
    Dim TestIDControl As Object  
  
    Set TestIDMenuBar = CommandBars.ActiveMenuBar  
    Set TestIDMenu = TestIDMenuBar.Controls.Add(Type:=msoControlPopup, _  
        Temporary:=True)  
    TestIDMenu.Caption = "1-300"  
    For I = 1 To 300  
        Set TestIDControl = TestIDMenu.Controls.Add(Type:=msoControlButton)  
        TestIDControl.FaceId = I  
        TestIDControl.Caption = Str(I)  
    Next I  
  
    Set TestIDMenuBar = CommandBars.ActiveMenuBar  
    Set TestIDMenu = TestIDMenuBar.Controls.Add(Type:=msoControlPopup, _  
        Temporary:=True)  
    TestIDMenu.Caption = "301-600"  
    For I = 301 To 600  
        Set TestIDControl = TestIDMenu.Controls.Add(Type:=msoControlButton)  
        TestIDControl.FaceId = I  
        TestIDControl.Caption = Str(I)  
    Next I  
End Sub
```

Рабочая книга, содержащая данный макрос, может медленно загружаться на компьютерах с низкой мощностью процессора. В этом случае, можно уменьшить значения счетчика I и создать, например, не 2 пункта меню: от 1 до 300 и от 301 до 600, как в показанном листинге, а один — от 1 до 100. Затем можно проверить, опять однократно, вариант от 101 до 200, и т. д.

Методы меню

Свойства `OnAction` в листинге 1.3 указывают имена методов, которые должны быть исполнены при выборе пользователем того или иного пункта меню. Это могут быть стандартные методы класса `Application` или специально составленные пользовательские функции, как показано на листинге 1.5.

Листинг 1.5

```
Private Sub MakeHelp()  
    Application.Help "Cla.hlp"  
End Sub  
  
Private Sub MakeAboutMain()  
    MainUserForm.Show  
End Sub  
  
Private Sub RunSolution()  
    ClaUserForm1.Show  
End Sub  
  
Private Sub MakeAbout()  
    ClaUserForm2.Show  
End Sub
```

Метод `Application.Help` заданием параметров вызывает соответствующую справочную систему, с указанием полного пути к тому каталогу, где находится файл справки. Программисту нужно заранее позаботиться о том, чтобы файл справки уже находился в указанном месте (или в каталоге по умолчанию) до того, как пользователь решит им воспользоваться. Метод одинаково хорошо работает и со справочной системой в формате HLP, и в формате HTML. О создании справочных систем см. *главу 12*. Там же подробно описано местонахождение файлов различных типов, применяемых в наших разработках.

Метод `MainUserForm.Show` вызывает форму (об их проектировании см. следующий раздел) с информацией о проекте `AtteStat`. Данная форма — общая для всех модулей, составляющих одну программу.

Метод `ClaUserForm1.Show` вызывает диалоговую панель, предназначенную для выбора параметров расчета и запуска прикладных алгоритмов (рис. 7.2). В нашем проекте, помимо общей формы с информацией о программном обеспечении `AtteStat`, каждый модуль имеет свою собственную информационную форму.

Метод `ClauUserForm2.Show` вызывает форму с информацией о модуле (рис. 1.3). Данная форма — специфическая для каждого разработанного в рамках нашего проекта модуля.

Все представленные в данном модуле функции описаны как `Private`, чтобы исключить их случайный вызов из листа рабочей книги или других модулей и исключить конфликты имен в случае, если в других модулях окажутся функции с такими же именами. Если функции специально предназначены для вызова за пределами содержащего их модуля, они должны быть описаны как `Public`.

Создание диалоговых панелей

Для обеспечения удобства пользователей и упрощения работы программиста, в VBA предусмотрена возможность быстрого создания диалоговых панелей (форм). На рис. 1.4 показано окно редактора Visual Basic в режиме проектирования формы. Неопытным в программировании читателям не следует пугаться слишком сложной картины. Во-первых, редактор весьма удобен. Во-вторых, сложность формы зависит от сложности решаемой задачи. В-третьих, на прилагаемом к книге компакт-диске содержатся готовые коды всех рассмотренных модулей, которые можно использовать в качестве основы для собственных разработок.

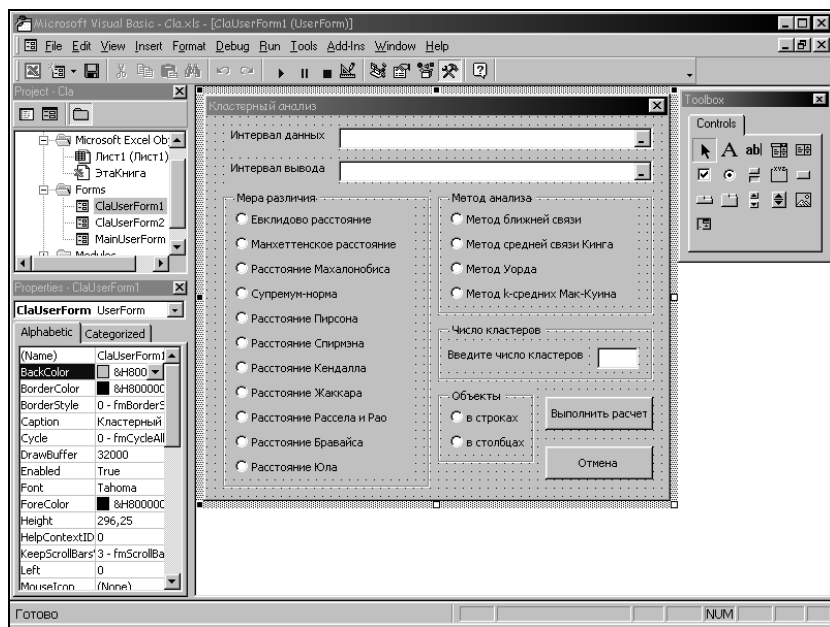


Рис. 1.4. Окно редактора Visual Basic в режиме проектирования формы